具有性能保证的分布式学习隐私保护

Elsa Rizk*，*IEEE* 会员，Stefan Vlaski†，*IEEE* 会员，以及 Ali H. Sayed*，*IEEE* 会士

arXiv:2301.06412v1 [cs.LG] 16 Jan 2023

摘要——**我们研究分布式学习与优化策略的隐私化，重点关注差分隐私方案并考察其对性能的影响。我们表明，广受欢迎的加性随机扰动方案会降低性能，因为它未能针对图结构进行恰当调优。因此，我们采用两种替代性的图同态构造，并表明它们在保障隐私的同时能够提升性能。此外，与大多数早期研究不同，我们不假设风险函数的梯度有界（这一条件在实践中很少成立，例如二次风险函数）。我们避开这一条件，仍然设计出一种以高概率实现差分隐私的方案。我们研究优化与学习场景，并通过仿真展示理论结果。**

索引词——**分布式学习，隐私化学习，差分隐私，分布式优化**

I. 引言

DISTRIBUTED 学习与优化策略适用于许多现实问题场景，例如为救援任务设计机器人群、设计云计算服务，或在社交网络中交换信息。即使在可以采用集中式解决方案的情况下，出于各种原因，通常也更倾向于采用分布式实现。例如，集中式解决方案往往维护成本高，并且容易受到中央处理器故障的影响。出于隐私和安全方面的考虑，智能体也可能不愿将数据分享给远程中央处理器。每个智能体可用的数据量可能非常庞大，因此难以定期在分散的智能体与中央处理器之间传输大量数据。分布式实现提供了一种极具吸引力且稳健的替代方案。由于处理工作可以由其余智能体继续完成，该架构能够容忍个别智能体发生故障。此外，智能体只需与邻居分享经过处理的少量信息。

分布式优化有多种方案，文献中对此已有广泛研究。这些方案包括增量策略 [1]–[8]，、共识策略 [9]–[18]，和扩散策略 [19]–[24]。增量算法需要沿循环路径对智能体重新编号，以遍历整个图。这通常是一项具有挑战性的任务，因为确定合适的循环路径是一个 NP难问题，而且路径上任何一条边发生故障，都会使该方案失去作用。

*作者任职于洛桑联邦理工学院工程学院（电子邮箱：{elsa.rizk; ali.sayed}@epfl.ch）。†作者任职于伦敦帝国理工学院电气与电子工程系（电子邮箱：s.vlaski@imperial.ac.uk）。

共识策略和扩散策略无需在图上构建循环路径。它们依赖相邻智能体之间的局部信息共享。这两类策略的一个主要区别在于，一致性更新是不对称的：梯度下降步的起点与梯度向量的评估位置不同——参见表达式 (16)。多项早期研究（例如，见 [20], [25], [26]）表明，与扩散方案相比，这种不对称性会缩小共识实现的稳定范围，尤其是在需要持续学习和自适应的场景中。

分布式架构的一个关键特点是，它们要求智能体与邻居共享信息。这一特点引出了一个重要的隐私问题：图中边上传输的信息是否可能被截获。例如，已知在依赖梯度下降更新的算法中，共享局部梯度或其估计的模型可能导致信息泄露 [27]–[30]。当网络处理医疗或金融数据等机密或敏感数据时，这可能会造成问题。在这种情况下，攻击者可能能够恢复个人个人信息中的某些内容。毫无疑问，采用能够保障一定隐私水平的分布式策略是有益的。

文献中已有多项有益研究探讨了分布式算法中的隐私问题。这些研究主要依赖两类工具：差分隐私或密码学。密码学方法包括安全聚合、多方计算和同态加密 [31]–[35]。尽管这些方法不会妨碍所学习模型的性能，但会带来显著的计算开销和通信开销。

另一方面，差分隐私方法通过添加一些随机噪声来掩盖消息 [36]–[46]。这些方法易于实现，但会给学习到的模型引入误差，并降低网络的整体效用。这种性能下降的一个主要原因是，添加噪声时往往没有考虑图拓扑。

本研究聚焦于差分隐私，因为它更易于应用，且可扩展性更强。我们阐述了如何根据图拓扑调整其应用方式，同时确保隐私和性能保证。具体而言，我们考察了两种差分隐私方案的效果：传统的随机扰动方案和图同态方案。我们证明，就均方误差（MSE）而言，后者优于前者。我们还设计了第三种方案，称为局部图同态处理，它能完全消除

噪声对性能的劣化影响。这些结果适用于广泛的一类分布式学习和优化形式。

II. 问题设定

我们考虑一个包含 P 个智能体的图拓扑，这些智能体标记为$p=1,\ 2,\ \ldots,\ P$，如图1所示。我们的目标是让这些智能体趋近于如下形式的聚合凸优化问题的极小化解：

$$w^o \triangleq \operatorname*{argmin}_{w\in\mathbb{R}^M} \frac{1}{P}\sum_{p=1}^{P}\left\{ J_p(w) \triangleq \frac{1}{N_p}\sum_{n=1}^{N_p} Q_p(w;x_{p,n})\right\}, \tag{1}$$

其中，风险函数 $J_p(\cdot)$与第 p个智能体相关，其定义为对应损失函数 $Q_p(\cdot;\cdot)$在本地数据 $\{x_{p,n}\}_{n=1}^{N_p}$上的经验平均。对于连接相邻智能体 m 和 p的边，我们赋予两个非负权重 a_{mp} 和 a_{pm}。在此记号下，a_{mp} 是智能体 p 用于缩放来自 m的信息的权重；同理，a_{pm}用于缩放从 p传向 m的信息。智能体 p 的邻域记为 $\mathcal{N}_p$，由所有通过边与 p 相连的智能体组成。我们假设 $\mathcal{N}_p$ 也包含智能体 p。

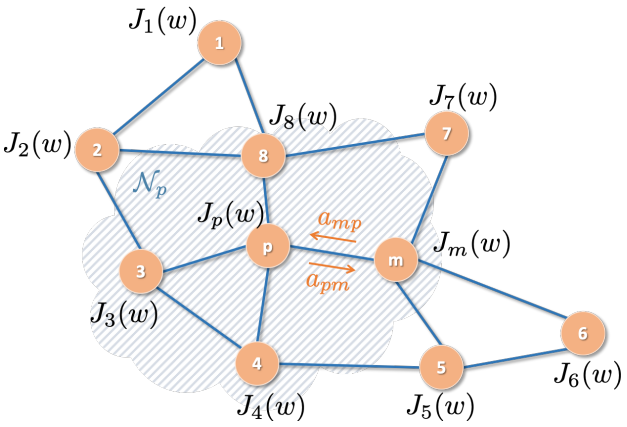


图1：智能体网络示意图。

A. 建模条件

我们假设各个风险函数 $J_p(w)$ 都是强凸的，损失函数 $Q_p(w;\cdot)$ 具有利普希茨连续的梯度且二阶可微。在分布式方法的研究中，这些条件十分常见。虽然这些条件可以放宽，所得结果也可以推广到更广泛的情形（例如，参见 [25], [26], [47]–[49]），但在本研究中，基于这些假设阐释主要思想已足够。

假设 1 (凸性与光滑性)。 风险函数 $J_p(\cdot)$是 ν–强凸的，而损失函数 $Q_p(\cdot;\cdot)$ 是凸且二阶可微的，即对于某个 $\nu>0$:

$$J_p(w_2) \geq J_p(w_1) + \nabla_{w^\mathsf{T}} J_p(w_1)(w_2-w_1) + \frac{\nu}{2}\|w_2-w_1\|^2, \tag{2}$$

$$Q_p(w_2;\cdot) \geq Q_p(w_1;\cdot) + \nabla_{w^\mathsf{T}} Q_p(w_1;\cdot)(w_2-w_1). \tag{3}$$

损失函数具有 δ–利普希茨连续梯度，这意味着存在 $\delta>0$，使得对于任意数据点 $x_{p,\ n}$:

$$\|\nabla_{w^\mathsf{T}} Q_p(w_2;x_{p,n}) - \nabla_{w^\mathsf{T}} Q_p(w_1;x_{p,n})\| \leq \delta\|w_2-w_1\|. \tag{4}$$

□

由于我们假设损失函数二阶可微，因此上述强凸性和利普希茨连续性条件等价于（参见 [25], [26], [49]）：

$$0 < \nu I \leq \nabla^2_{w^\mathsf{T}} J_p(w) \leq \delta I. \tag{5}$$

我们进一步假设图拓扑是强连通的。这意味着任意一对智能体（$m,\ p$）之间都存在双向连接的路径，而且网络中至少有一个智能体 p 带有一个取值为 $a_{pp}>0$ 的自环。换言之，至少有一个智能体信任其本地信息。组合矩阵 $A=[a_{mp}]$ 通常是左随机的，也就是说，其元素满足：

$$a_{mp} \geq 0, \quad \sum_{m\in\mathcal{N}_p} a_{mp} = 1. \tag{6}$$

也就是说，连接智能体的边上的权重非负，且 A 的每一列中的元素之和为 1。图的强连通性意味着 A 是一个本原矩阵。因此，根据佩龙-弗罗贝尼乌斯定理 [26]，A 将有一个等于 1 的单一特征值，而所有其他特征值都严格位于单位圆内。此外，存在一个特征向量 q，其元素均为正 $\{q_p\}$，且元素之和为 1，并满足：

$$Aq = q, \quad q_p>0, \quad \mathbb{1}^\mathsf{T} q = 1. \tag{7}$$

我们将 q 称为 A 的 Peron 特征向量。此外，有 $\rho\left(A-q\mathbb{1}^\mathsf{T}\right)<1$，其中 $\rho(\cdot)$ 表示其矩阵参数的谱半径。

接下来，令 w^o 表示式 (1) 的全局极小化解，并令 w_p^o 表示 $J_p(\cdot)$ 的局部极小化解。我们假设这些全局模型与局部模型之间的差异有界；否则，协作将不会带来好处，人们会改用其他优化方法，例如多任务学习 [50]。

为了进一步阐明这一点，我们考虑一个涉及二次损失的简单示例。假设到达节点p的数据（记为 $\boldsymbol{d}_p(n)$）由某个线性回归模型在加性噪声作用下生成，其形式如下：

$$\boldsymbol{d}_p(n) = \boldsymbol{u}_{p,n}^\mathsf{T} w^\star + \boldsymbol{o}_p(n), \tag{8}$$

其中 $\boldsymbol{u}_{p,n}$ 是特征向量，$w^\star$ 是模型。我们可以通过求解以下问题来估计 $w^\star$：

$$\min_w \frac{1}{P}\sum_{p=1}^{P}\frac{1}{N_p}\sum_{n=1}^{N_p}(\boldsymbol{d}_p(n) - \boldsymbol{u}_{p,n}^\mathsf{T} w)^2. \tag{9}$$

此时的全局极小化解为：

$$w^o = w^\star + \widehat{R}_u^{-1}\widehat{r}_{uo}, \tag{10}$$

其中:

$$\widehat{R}_u \triangleq \frac{1}{P}\sum_{p=1}^{P}\left\{\widehat{R}_{p,u} \triangleq \frac{1}{N_p}\sum_{n=1}^{N_p}\boldsymbol{u}_{p,n}\boldsymbol{u}_{p,n}^{\mathsf{T}}\right\}, \tag{11}$$

$$\widehat{r}_{uo} \triangleq \frac{1}{P}\sum_{p=1}^{P}\left\{\widehat{r}_{p,uo} \triangleq \frac{1}{N_p}\sum_{n=1}^{N_p}\boldsymbol{o}_p(n)\boldsymbol{u}_{p,n}\right\}, \tag{12}$$

而 $J_p(w)$的局部极小值为:

$$w_p^o = w^\star + \widehat{R}_{p,u}^{-1}\widehat{r}_{p,uo}. \tag{13}$$

因此，全局模型(10)可以表示为局部模型(13)的加权平均:

$$w^o = \frac{1}{P}\sum_{p=1}^{P}\widehat{R}_u^{-1}\widehat{R}_{p,u}w_p^o. \tag{14}$$

这意味着全局模型是局部模型的混合。因此，下面对模型差异施加的界相当于对各智能体之间数据分布差异程度的假设。该条件比对代价函数的梯度之差施加统一上界更宽松，而后者在文献中更为常见（参见 [51], [52]）。

假设2 (模型漂移)。 每个局部模型 w_p^o 与全局模型 w^o 之间的距离均有界，即存在 $\xi \geq 0$ 使得 $\|w^o - w_p^o\| \leq \xi$。

□

III. 分布式学习

A. 广义分布式学习

我们重点关注两种主要策略：共识和扩散。用于求解 (1) 的一致性策略形式如下:

$$\boldsymbol{\psi}_{p,i-1} = \sum_{m\in\mathcal{N}_p} a_{mp}\boldsymbol{w}_{m,i-1}, \tag{15}$$

$$\boldsymbol{w}_{p,i} = \boldsymbol{\psi}_{p,i-1} - \mu\widehat{\nabla_{w^{\mathsf{T}}}J_p}(\boldsymbol{w}_{p,i-1}), \tag{16}$$

其中，$\widehat{\nabla_{w^{\mathsf{T}}}J_p}(\cdot)$ 表示对 $J_p(\cdot)$ 的真实梯度的随机梯度近似。通常，近似值取为损失函数的梯度，即 $\nabla_{w^{\mathsf{T}}}Q_p(\boldsymbol{w}_{p,\,i-1},\,\boldsymbol{x}_{p,\,i})$。这里，量 $\{\boldsymbol{\psi}_{p,i}, \boldsymbol{w}_{p,i}\}$ 表示节点 p 在时间 i对 w^o 的估计值。注意，(16) 中的梯度向量是在先前的本地模型 $\boldsymbol{w}_{p,\,i-1}$ 处计算的，而不是在中间模型 $\boldsymbol{\psi}_{p,\,i-1}$处计算的。扩散策略则有两种相关的实现方式，分别称为先组合后适应（CTA）和先适应后组合（ATC）。它们的区别在于计算顺序：一种情况下先组合后自适应，另一种情况下顺序相反。CTA 扩散策略描述如下:

$$\boldsymbol{\psi}_{p,i-1} = \sum_{m\in\mathcal{N}_p} a_{mp}\boldsymbol{w}_{m,i-1}, \tag{17}$$

$$\boldsymbol{w}_{p,i} = \boldsymbol{\psi}_{p,i-1} - \mu\widehat{\nabla_{w^{\mathsf{T}}}J_p}(\boldsymbol{\psi}_{p,i-1}). \tag{18}$$

与(15)–(16)比较，现在可以看到，式(18)中梯度下降步的起点与以下点相同

梯度向量就在该点处求值。同理，ATC 扩散策略如下:

$$\boldsymbol{\psi}_{p,i} = \boldsymbol{w}_{p,i-1} - \mu\widehat{\nabla_{w^{\mathsf{T}}}J_p}(\boldsymbol{w}_{p,i-1}), \tag{19}$$

$$\boldsymbol{w}_{p,i} = \sum_{m\in\mathcal{N}_p} a_{mp}\boldsymbol{\psi}_{m,i}. \tag{20}$$

上述三种算法可以组合成如下的统一描述 [26]:

$$\boldsymbol{\phi}_{p,i-1} = \sum_{m\in\mathcal{N}_p} a_{1,mp}\boldsymbol{w}_{m,i-1}, \tag{21}$$

$$\boldsymbol{\psi}_{p,i} = \sum_{m\in\mathcal{N}_p} a_{0,mp}\boldsymbol{\phi}_{m,i-1} - \mu\widehat{\nabla_{w^{\mathsf{T}}}J_p}(\boldsymbol{\phi}_{p,i-1}), \tag{22}$$

$$\boldsymbol{w}_{p,i} = \sum_{m\in\mathcal{N}_p} a_{2,mp}\boldsymbol{\psi}_{m,i}. \tag{23}$$

其中，我们引入了三个组合矩阵$\{A_0, A_1, A_2\}$。令 $A_0 = A$ 和 $A_1 = A_2 = I$，即可得到共识；令 $A_1 = A$ 和 $A_0 = A_2 = I$，则得到 CTA 扩散；令 $A_2 = A$ 和 $A_0 = A_1 = I$，则得到 ATC 扩散。也可以采用其他选择。

B. 隐私学习

现在，我们来考察差分隐私算法，以保护智能体之间共享信息的隐私。为便于说明，假设智能体 1 处的数据 $\{x_{1,n}\}$ 被另一组数据 $\{x'_{1,n}\}$ 替代。因此，该算法将产生一条新的轨迹，记为 $\{\boldsymbol{\phi}'_{p,i-1}, \boldsymbol{\psi}'_{p,i}, \boldsymbol{w}'_{p,i}\}$。在隐私实现中，外部观察者不应察觉智能体 1 处的这一变化。具体来说，我们只需对需要进行隐私化的消息添加噪声。最常见的做法是添加服从指数分布的噪声，例如拉普拉斯分布或高斯分布噪声 [36]。因此，我们首先考虑如下形式的隐私化分布式实现:

$$\boldsymbol{\phi}_{p,i-1} = \sum_{m\in\mathcal{N}_p} a_{1,mp}\left(\boldsymbol{w}_{m,i-1} + \boldsymbol{g}_{1,mp,i}\right), \tag{24}$$

$$\boldsymbol{\psi}_{p,i} = \sum_{m\in\mathcal{N}_p} a_{0,mp}\left(\boldsymbol{\phi}_{m,i-1} + \boldsymbol{g}_{0,mp,i}\right) - \mu\widehat{\nabla_{w^{\mathsf{T}}}J_p}(\boldsymbol{\phi}_{p,i-1}), \tag{25}$$

$$\boldsymbol{w}_{p,i} = \sum_{m\in\mathcal{N}_p} a_{2,mp}\left(\boldsymbol{\psi}_{m,i} + \boldsymbol{g}_{2,mp,i}\right), \tag{26}$$

其中，$\boldsymbol{g}_{j,\,mp,\,i}$ 表示对于每个 $j = 0$、1、2，m，$p = 1$、2、$\cdots$、P，均为均值为零的拉普拉斯随机噪声。例如，在（24）中，智能体 m 通过连接双方的边，与智能体 p 共享 $\boldsymbol{w}_{m,\,i-1}$。在此传输过程中，会添加大小为$\boldsymbol{g}_{1,\,mp,\,i}$ 的拉普拉斯噪声。下标 mp 用于表示该噪声对应从 m 到p的有向通信。其他噪声也类似。

接下来，我们正式定义差分隐私 [36], ，并证明上述算法确实满足差分隐私。

定义 1 ($\epsilon(i)$–**差分隐私**)。 我们称由 (24)–(26) 给出的算法，对于智能体 p 在时间 i 是 $\epsilon(i)$–差分隐私的，当且仅当针对观察相应事件的概率满足以下条件:

联合分布 $f(\cdot)$ 其中，记号 $\boldsymbol{y}_{p,\ j-1}$ 表示任意共享消息 $\{\boldsymbol{w}_{p,j-1}, \boldsymbol{\phi}_{p,j}, \boldsymbol{\psi}_{p,j}\}$, $\boldsymbol{g}_{\cdot,\ pm,\ j}$ 及其对应的添加噪声 $\{\boldsymbol{g}_{1,pm,j}, \boldsymbol{g}_{0,pm,j}, \boldsymbol{g}_{2,pm,j}\}$:

$$\frac{f\left(\left\{\{\boldsymbol{y}_{p,j-1}+\boldsymbol{g}_{\cdot,pm,j}\}_{m\in\mathcal{N}_p\setminus\{p\}}\right\}_{j=1}^{i}\right)}{f\left(\left\{\{\boldsymbol{y}'_{p,j-1}+\boldsymbol{g}_{\cdot,pm,j}\}_{m\in\mathcal{N}_p\setminus\{p\}}\right\}_{j=1}^{i}\right)}\leq e^{\epsilon(i)}. \tag{27}$$

□

上述界限保证，当 $\epsilon(i)$ 较小时，原始轨迹与修改后轨迹的分布彼此接近。因此，由于我们无法区分参与智能体组合不同时算法产生的轨迹，便难以推断智能体所持数据的相关信息。换言之，如果代理 1 选择用 $\{x'_{1,n}\}$ 替换其原始数据集，那么由此产生的模型 $\{\boldsymbol{w}'_{p,j-1}, \boldsymbol{\phi}'_{p,j}, \boldsymbol{\psi}'_{p,j}\}$ 与原始模型 $\{\boldsymbol{w}_{p,j-1}, \boldsymbol{\phi}_{p,j}, \boldsymbol{\psi}_{p,j}\}$ 的分布足够接近，外部观察者将无法判断使用的是哪个数据集。使用原始数据集和替代数据集所产生的两条模型轨迹无法区分。

为证明算法 (24)–(26) 满足条件 (27)，我们首先计算该算法的敏感度。附录A 将时刻 i 的敏感度定义为：若代理 1 改用替代数据集 $\{x'_{1,n}\}$，而非使用原始数据集，算法轨迹所发生的变化。附录A 中证明，该敏感度满足:

$$\Delta(i)\ \triangleq\ \|\boldsymbol{\mathcal{W}}_i-\boldsymbol{\mathcal{W}}'_i\|\leq B+B'+\sqrt{P}\|w^o-w'^o\|, \tag{28}$$

对于某些常数 B 和 B'，且以高概率成立。也就是说，有:

$$\begin{aligned}&\mathbb{P}(\Delta(i)\leq B+B'+\sqrt{P}\|w^o-w'^o\|)\\&\geq\left(1-\frac{\kappa_2^2\mathbb{1}^\mathsf{T}\Gamma^i\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0\|^2\\ \mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_0\|^2\end{bmatrix}+O(\mu)+O(\mu^{-1})}{B^2}\right)\\&\times\left(1-\frac{\kappa_2'^2\mathbb{1}^\mathsf{T}\Gamma'^i\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}'_0\|^2\\ \mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}'_0\|^2\end{bmatrix}+O(\mu)+O(\mu^{-1})}{B'^2}\right),\end{aligned} \tag{29}$$

其中，$\boldsymbol{\mathcal{W}}_i\ \triangleq\ \mathrm{col}\{\boldsymbol{w}_{p,i}\}_{p=1}^P$表示，零时刻的模型误差记为 $\tilde{\boldsymbol{\mathcal{W}}}_0\ \triangleq\mathrm{col}\{w^o-\boldsymbol{w}_{p,0}\}_{p=1}^P$，变量 $\{\bar{\boldsymbol{\mathcal{W}}}_0, \check{\boldsymbol{\mathcal{W}}}_0\}$ 源自对 $\mathcal{V}_\theta^\mathsf{T}\tilde{\mathcal{W}}_0=\mathrm{col}\{\bar{\boldsymbol{\mathcal{W}}}_0, \check{\boldsymbol{\mathcal{W}}}_0\}$的划分，其中使用了矩阵 Γ 和附录 B 中定义的常数 κ_2。结果（29）意味着，敏感度 $\Delta(i)$ 以高概率有界。界常数 B 和 B' 由用户选择: B 和 B' 取值越大，敏感度有界的概率越高；但如（31）所示，这会导致更大的隐私界。换言之，可以控制 B 和 B'的取值，以权衡隐私级别与敏感度有界的可能性。接下来，若将拉普拉斯噪声 $\boldsymbol{g}_{j,\ mp,\ i}$ 的方差记为 σ_g^2,

其中 $\boldsymbol{y}_{p,\ j-1}=\boldsymbol{w}_{p,\ j-1}$，（27）中的分式以高概率可按如下方式进行界定:

$$\begin{aligned}&\frac{f\left(\left\{\{\boldsymbol{w}_{p,j-1}+\boldsymbol{g}_{0,pm,j}\}_{m\in\mathcal{N}_p\setminus\{p\}}\right\}_{j=1}^{i}\right)}{f\left(\left\{\{\boldsymbol{w}'_{p,j-1}+\boldsymbol{g}_{0,pm,j}\}_{m\in\mathcal{N}_p\setminus\{p\}}\right\}_{j=1}^{i}\right)}\\&\overset{(a)}{=}\prod_{j=1}^{i}\frac{f\left(\{\boldsymbol{w}_{p,j-1}+\boldsymbol{g}_{0,pm,j}\}_{m\in\mathcal{N}_p\setminus\{p\}}|\mathcal{X}_{j-1}\right)}{f\left(\{\boldsymbol{w}'_{p,j-1}+\boldsymbol{g}_{0,pm,j}\}_{m\in\mathcal{N}_p\setminus\{p\}}|\mathcal{X}'_{j-1}\right)}\\&\overset{(b)}{=}\prod_{j=1,m\in\mathcal{N}_p\setminus\{p\}}^{i}\frac{\exp\left(-\sqrt{2}\|\boldsymbol{w}_{p,j-1}+\boldsymbol{g}_{0,pm,j}\|/\sigma_g\right)}{\exp\left(-\sqrt{2}\|\boldsymbol{w}'_{p,j-1}+\boldsymbol{g}_{0,pm,j}\|/\sigma_g\right)}\\&\leq\exp\left(\frac{\sqrt{2}}{\sigma_g}\sum_{j=1,m\in\mathcal{N}_p\setminus\{p\}}^{i}\|\boldsymbol{w}_{p,j-1}-\boldsymbol{w}'_{p,j-1}\|\right)\\&\leq\exp\left(\frac{\sqrt{2}P}{\sigma_g}\sum_{j=1}^{i}\|\boldsymbol{\mathcal{W}}_{j-1}-\boldsymbol{\mathcal{W}}'_{j-1}\|\right),\end{aligned} \tag{30}$$

其中，第一个等式（a）由对 $\mathcal{X}_{j-1}\triangleq\{\boldsymbol{w}_{p,j-1}\}\cup\left\{\{\boldsymbol{w}_{p,o-1}+\boldsymbol{g}_{0,pm,o}\}_{m\in\mathcal{N}_p\setminus\{p\}}\right\}_{o=1}^{j-1}$ 应用贝叶斯定理得到，第二个等式（b）则由在给定 $\boldsymbol{w}_{p,j-1}$ 的条件下，$\boldsymbol{w}_{p,j-1}+\boldsymbol{g}_{0,\ pm,j}$ 对于 $m\in\mathcal{N}_p\setminus\{p\}$ 的独立性得到。对于 $\boldsymbol{y}_{p,j-1}\in\{\boldsymbol{\phi}_{p,j}, \boldsymbol{\psi}_{p,j}\}$ 也可以得到类似的界。

因此，隐私级别由以下选择来定义：根据运行中的 $\Delta(j)$ 值确定 $\epsilon(i)$:

$$\epsilon(i)=\frac{\sqrt{2}P}{\sigma_g}\sum_{j=0}^{i-1}\Delta(j)\leq\frac{\sqrt{2}P}{\sigma_g}(B+B'+\sqrt{P}\|w^o-w'^o\|)i. \tag{31}$$

这些结果表明，要得到一个$\epsilon(i)$–差分隐私算法，只需选择拉普拉斯噪声的方差，使其满足（31）。表达式（31）表明，$\epsilon(i)$ 是迭代次数的线性函数。这意味着，该过程的隐私性降低速率不会超过线性速率。需要指出的是，关于多智能体系统差分隐私方案的大多数早期研究 [38], [41], [42] 都假设风险函数的梯度有界。然而，这一条件在实践中很少成立。例如，即使是二次风险函数，其梯度也可能无界。因此，我们的方法避免依赖这一假设。相反，我们能够证明，差分隐私仍以高概率成立。

我们仍需考察添加的噪声对性能的影响。为此，我们引入扩展模型$\boldsymbol{\mathcal{W}}_i\ \triangleq\ \mathrm{col}\{\boldsymbol{w}_{p,i}\}_{p=1}^P$，并将三步算法 (24)–(26) 写成如下的单次递推:

$$\begin{aligned}\boldsymbol{\mathcal{W}}_i=&\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{W}}_{i-1}-\mu\mathcal{A}_2^\mathsf{T}\mathrm{col}\left\{\widehat{\nabla_{w^\mathsf{T}}J_p}(\boldsymbol{\phi}_{p,i-1})\right\}_{p=1}^P\\&+\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})+\mathcal{A}_2^\mathsf{T}\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i})\\&+\mathrm{diag}(\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i}),\end{aligned} \tag{32}$$

其中，对于 $j=0,\ 1,\ 2$, $\mathcal{A}_j\ \triangleq\ A_j\otimes I_M$ 和 $\boldsymbol{\mathcal{G}}_{j,\ i}$ 是一个矩阵，其元素为添加的噪声 $\boldsymbol{g}_{j,\ mp,\ i}$。我们将

模型误差记为 $\widetilde{\boldsymbol{\mathcal{W}}}_i \triangleq \mathrm{col}\{w^o - \boldsymbol{w}_{p,i}\}_{p=1}^P$，并引入局部梯度噪声：

$$\boldsymbol{s}_{p,i} \triangleq \widehat{\nabla_{w^\mathsf{T}} J_p}(\boldsymbol{\phi}_{p,i-1}) - \nabla_{w^\mathsf{T}} J_p(\boldsymbol{\phi}_{p,i-1}). \tag{33}$$

通常假设该梯度噪声过程均值为零且二阶矩有界（例如，参见[26], [49], ，其中实际上证明了这一性质在许多重要情形下均成立，类似的论证也可用于当前情形），即：

$$\mathbb{E}\{\|\boldsymbol{s}_{p,i}\|^2|\mathcal{F}_{i-1}\} \le \beta_{s,p}^2\|\widetilde{\boldsymbol{\phi}}_{p,i-1}\|^2 + \sigma_{s,p}^2, \tag{34}$$

对于某些非负常数 $\beta_{s,\,p}^2$ 和 $\sigma_{s,\,p}^2$，其中条件期望是针对所有过去的模型 $\mathcal{F}_{i-1} \triangleq$ 滤过$\{\boldsymbol{w}_{p,j}\}^{P,\ p=1,\ j=0}_{\ \ \ \

引理 1 (网络分歧)。 每次迭代 i 中，偏离质心的平均值均有统一上界；此外，当 i 趋于无穷时，有：

$$\limsup_{i\to\infty}\frac{1}{P}\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{w}_{p,i}-\boldsymbol{w}_{c,i}\|^2\le O(1)\sigma_g^2+O(\mu) \quad (47)$$

证明。 参见附录 C。 □

表达式（47）表明，局部模型与质心模型之间的距离至多为$O(1)\sigma_g^2$。因此，如果质心模型能够以微小偏差收敛到最优模型 w^o，那么局部模型与真实模型之间的距离始终为一个与噪声方差 σ_g^2成正比的常数。在下一个定理中，我们将证明，添加噪声对质心模型的影响仅为 $O(1)$因子。

定理 2 (网络质心的均方误差收敛)。在假设 *1* 和 *2* 下，由（45）定义的网络质心在步长足够小时，以指数速度收敛到最优模型的一个邻域：

$$\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i}\|^2\le O(\mu)\sigma_s^2+O(1)\sigma_g^2+O(\mu^2). \quad (48)$$

证明。 参见附录 D。 □

因此，即使存在添加噪声，网络质心与真实极小值点 w^o之间的距离也至多为 $O(1)\sigma_g^2$，而非 $O(\mu^{-1})\sigma_g^2$。在引理 1 中，我们证明了个体模型 $\boldsymbol{w}_{p,\,i}$ 与质心模型之间的距离为 $O(1)\sigma_g^2$。因此，采用图同态扰动（43）–（44）后，均方误差不再与 μ 成反比，相较于（42）有所改进。

D. 局部图同态噪声

我们将说明如何改进 $O(1)\sigma_g^2$ 偏差，并通过使用局部图同态噪声将其替换为 $O(\mu)\sigma_g^2$。为此，我们构造噪声，使其满足以下替代条件，而非条件 (43)：

$$\sum_{m\in\mathcal{N}_p}a_{mp}\boldsymbol{g}_{j,mp,i}=0. \quad (49)$$

请注意，我们要求噪声之和在局部相互抵消，而不是像上一节所要求的那样在全局范围内抵消。每个智能体 p 的邻居必须协同生成相关随机噪声，使其在 p 处局部抵消。由于这些邻居之间可能不存在直接连接，因此协作将通过智能体 p 进行。区块链应用中也存在类似问题，即需要以分布式方式生成随机数 [53]。

现在，我们设计一种分布式方案，使其产生的噪声满足条件 (49)。为便于说明，我们通过一个示例来描述该协议。假设智能体 1 与 5 个智能体相连，分别标记为 2、3、4、5、6（图 2，左侧）。由于智能体 1 的邻居之间不一定通过直接链路彼此相连，因此所有通信都将通过智能体 1 进行。在这个子网络中，我们让智能体 1 担任该方案的协调者。第一步是由智能体 1 将其邻居划分为两个不相交集合 $\mathcal{N}_1=\mathcal{N}_+\bigcup\mathcal{N}_-$。

例如，我们可以将编号为偶数的智能体归入$\mathcal{N}_+$，将编号为奇数的智能体归入 $\mathcal{N}_-$。然后，我们让来自两个不相交集合的每一对智能体协商一个噪声值，并将其添加到各自的消息中，使这些噪声值在智能体 1 处相互抵消。我们要求来自 $\mathcal{N}_-$ 的智能体在添加到消息中的噪声前乘以负号。因此，智能体 2 会在其消息中添加两个噪声项，一个由其与智能体 3 生成，另一个由其与智能体 5 生成。我们用 $\boldsymbol{g}_{\{23\}1,\,i}$ 表示由智能体 2 和智能体 3 生成并将发送给智能体 1 的噪声项。由于消息会按智能体 1 分配给其邻居的权重进行缩放，因此添加噪声时必须除以相应权重，即智能体 2 发送给智能体 1 的消息由原始消息 $\boldsymbol{w}_{2,\,i}$ 以及智能体 2 与智能体 3、5 生成的两个噪声按相应权重缩放后组成 a_{12}：

$$\boldsymbol{w}_{2,i}+\frac{\boldsymbol{g}_{\{23\}1,i}}{a_{12}}+\frac{\boldsymbol{g}_{\{25\}1,i}}{a_{12}}. \quad (50)$$

然而，这要求代理 2 知道代理 1 分配给其消息的权重。因此，如果使用非双随机组合矩阵，代理 1 必须公开权重。代理 4 与代理 6 之间也会发生同样的过程，二者均与代理 3 和代理 5 进行交互。随后，发送给代理 1 的聚合消息最终将成为未掩码权重之和：

$$\sum_{k\in\mathcal{N}_+}a_{1k}\left(\boldsymbol{w}_{k,i}+\sum_{\ell\in\mathcal{N}_-}\frac{\boldsymbol{g}_{\{k\ell\}1,i}}{a_{1k}}\right)+\sum_{k\in\mathcal{N}_-}a_{1k}\left(\boldsymbol{w}_{k,i}-\sum_{\ell\in\mathcal{N}_+}\frac{\boldsymbol{g}_{\{\ell k\}1,i}}{a_{1k}}\right)=\sum_{k\in\mathcal{N}_1}a_{1k}\boldsymbol{w}_{k,i}. \quad (51)$$

下面介绍生成成对噪声项 $\boldsymbol{g}_{\{k\ell\}1,\,i}$的方法。我们采用 Diffie-Helman 密钥交换协议，其中每对智能体共享一个密钥，并用它来生成添加噪声。给定两个智能体，例如 2 和 3，我们假设它们各自的密钥分别为 v_2 和 v_3。智能体之间约定一个已知的模数 π 和底数 b。随后，智能体 2 广播其公钥 $V_2=(b^{v_2}\bmod\pi)$，智能体 3 也进行相同操作 $V_3=(b^{v_3}\bmod\pi)$。接着，智能体 2 计算 $v_{23}=(V_3^{v_2}\bmod\pi)=(b^{v_2v_3}\bmod\pi)$，其结果与智能体 3 计算的$v_{23}=(V_2^{v_3}\bmod p)=(b^{v_3v_2}\bmod\pi)$相同。因此，这两个智能体现在共享一个只有它们知道的密钥 v_{23}。然后可以将此密钥用作添加噪声，以掩盖消息，即$\boldsymbol{g}_{\{23\}1,\,i}=v_{23}$。然而，要使这一过程满足差分隐私，我们需要最终的添加噪声服从拉普拉斯分布 Lap(0，$\sigma_g/\sqrt{2}$)。接下来，我们将介绍一种称为局部图同态处理方案的方法，它能确保添加噪声服从拉普拉斯分布。图 2 右侧的子图展示了这一过程。

定义 2 (局部图同态过程)。 给定智能体 k 的一个子网络，以及邻居 $\ell\in\mathcal{N}_+$ 和$m\in\mathcal{N}_-$。令智能体 ℓ 从 $[0,1]$，上的均匀分布中采样两个密钥 $\boldsymbol{v}_\ell$ 和 $\boldsymbol{v}'_\ell$，并令智能体 m 从伽马分布 $\Gamma(2,1)$中采样其密钥 $\boldsymbol{v}_m$ 和 $\boldsymbol{v}'_m$。令 π 为某个较大的素数，并令 a 为 π的倍数。那么，对于：

$$\boldsymbol{v}_{\ell m}=a\,e^{-\boldsymbol{v}_\ell\boldsymbol{v}_m}\mod\pi, \quad (52)$$

$$\boldsymbol{v}'_{\ell m}=a\,e^{-\boldsymbol{v}'_\ell\boldsymbol{v}'_m}\mod\pi, \quad (53)$$

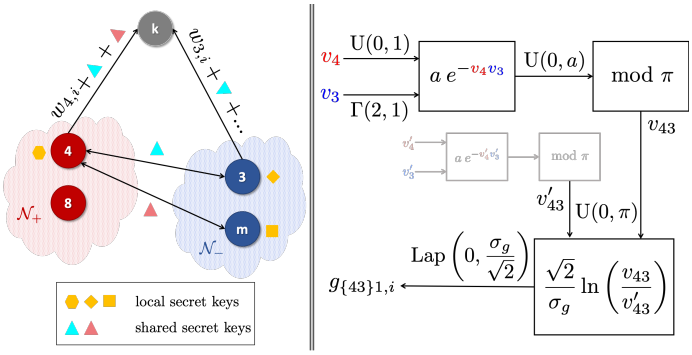


图 2：局部图同态过程示意图。左图描述了 Diffie-Hellman 密钥交换过程。右图展示了随机变量经历的变换。

所需的拉普拉斯噪声可构造如下：

$$\boldsymbol{g}_{\{\ell m\}k,i} = \frac{\sqrt{2}}{\sigma_g} \ln\left(\frac{\boldsymbol{v}_{\ell m}}{\boldsymbol{v}'_{\ell m}}\right). \tag{54}$$

□

本文提出的局部图同态过程与安全聚合领域的方法相关（见[31]）。然而，我们的方法与早期研究的主要区别在于，我们为更一般的分布式场景设计了一种方案，而其他研究主要关注联邦学习这一特例，其结构中包含一个中央处理器。此外，我们生成随机数以增强方案的安全性，而文献 [31] 则将伪随机数添加到共享消息中。由于伪随机数由确定性算法生成，因此会使噪声变得可预测且易受攻击，这与随机数相反。此外，与[31]不同，我们对方案的隐私性进行了量化。在下一个定理中，我们将证明，使用构造式（54）可得到一个差分隐私算法。

定理3 (采用局部图同态扰动的分布式学习的隐私性)。 在 (54) 定义的局部图同态过程中，所得到的隐私化算法以高概率满足 $\epsilon(i)$–差分隐私，其中 $\epsilon(i)$定义于(31)。

证明。 见附录E。 □

注意，由于每次迭代期间噪声在局部相互抵消，该算法遵循与非隐私化算法相同的轨迹。这意味着，采用局部图同态扰动(54)的隐私化分布式学习算法，其均方偏差性能与非隐私化分布式学习算法相同。具体而言， [26] 中定理 9.1 关于非隐私化算法收敛性的结果仍然成立。这意味着，现在均方误差将处于 $O(\mu)\sigma_g^2$的量级。

上述结果凸显了差分隐私面临的一项难题。请注意，原则上，随着添加噪声的方差 σ_g^2 增大，隐私级别也会提高。然而，这一过程会带来额外的通信成本。例如，代理 1 需要向其邻域传达如何划分正智能体和负智能体。它还需要与其邻居中近乎一半的智能体通信，以就添加噪声 $\boldsymbol{g}_{\{1m\}k,\ i}$ 达成一致，用于 $k \in \mathcal{N}_1$ 和 $m \in \mathcal{N}_k$。因此，代理 1 的通信成本总计最多会增加 $|\mathcal{N}_1| + \sum_{k=2}^{6} |\mathcal{N}_k|/2$。尽管额外的通信成本显然会影响隐私级别[1]，隐私度量却没有将其纳入考量。这降低了该度量的实用性，并促使人们寻求其他隐私指标，例如基于信息论度量的指标 [55]。例如，一种指标可以是原始消息与扰动后的共享消息之间的互信息 [54]。因此，假设各条消息 $\{\boldsymbol{w}_{k,i}\}$ 是方差为 $\sigma_{\boldsymbol{w}}^2$的高斯随机变量，并且我们用方差为 σ_g^2的总高斯噪声 $\boldsymbol{g}_{k,\ i}$ 对它们进行扰动，那么互信息为：

$$\begin{aligned} I(\boldsymbol{w}_{k,i}; \boldsymbol{w}_{k,i} + \boldsymbol{g}_i) &= H(\boldsymbol{w}_{k,i} + \boldsymbol{g}_{k,i}) - H(\boldsymbol{w}_{k,i} + \boldsymbol{g}_{k,i} | \boldsymbol{w}_{k,i}) \\ &= H(\boldsymbol{w}_{k,i} + \boldsymbol{g}_{k,i}) - H(\boldsymbol{g}_{k,i}) \\ &= \frac{1}{2} \log\left(1 + \frac{\sigma_{\boldsymbol{w}}^2}{\sigma_g^2}\right). \end{aligned} \tag{55}$$

再次注意，随着噪声方差 σ_g^2增大，互信息会降低，而隐私性会增强。互信息仍无法反映该过程产生的通信成本。目前文献中似乎还不存在能够刻画通信-隐私权衡的度量。因此，有必要为安全聚合方法寻找一种更合适的隐私度量。

IV. 实验分析

我们进行了两项实验。第一项实验聚焦于使用模拟数据的线性回归问题。随后，我们研究基于真实数据的分类问题。

A. 广义分布式隐私学习

对于共识、CTA 和 ATC 扩散，我们比较了四种算法：标准分布式算法、采用随机扰动的隐私化算法、采用图同态扰动的隐私化算法，以及采用局部图同态扰动的隐私化算法。我们考虑一个由 30 个智能体组成的网络（图3）和一个正则化二次损失函数：

$$\min_{w \in \mathbb{R}^2} \frac{1}{30} \sum_{p=1}^{30} \frac{1}{100} \sum_{n=1}^{100} (\boldsymbol{d}_p(n) - \boldsymbol{u}_{p,n}^{\mathsf{T}} w)^2 + 0.01\|w\|^2. \tag{56}$$

我们按如下方式生成随机数据集 $\{\boldsymbol{u}_{p,n}, \boldsymbol{d}_p(n)\}_{n=1}^{100}$ ：令二维特征向量 $\boldsymbol{u}_{p,n} \sim \mathcal{N}(0;\ R_u)$，并添加噪声 $\boldsymbol{o}_{p,n} = \boldsymbol{u}_{p,n}^{\mathsf{T}} w^\star + \boldsymbol{o}_{p,n}$，$w^\star \in \mathbb{R}^2$ $(\ ) \sim \mathcal{N}(0; \sigma_{o,}^2)$，使得 $\boldsymbol{d}_p(\ ),(\ )$，其中生成模型为某个模型，并随机设定方差 R_u 和添加噪声方差 $\sigma_{o,p}^2$。为使数据分布呈非独立同分布，我们使用不同的噪声

[1]如果减少局部图同态过程生成随机噪声的次数，转而重复使用噪声，那么通信成本会降低，但攻击者学习到噪声并解除消息掩码的可能性会增加。

方差$\sigma_{o,\ p}^2$ 在各个智能体之间。最优全局模型具有一个闭式解，其中 R_u 和 $\widehat{r}_{uo}$ 的定义同前：

$$w^o = (\widehat{R}_u + 0.01I)^{-1}(\widehat{R}_u w^\star + \widehat{r}_{uo}). \tag{57}$$

我们将步长设为 $\mu = 0.4$，噪声方差设为 σ_g^2 =0.01，并运行算法 1000 次迭代。我们重复该实验 20 次，并在对数域中绘制质心模型的均方偏差以及个体均方偏差的平均值：

$$\text{MSD}_i \triangleq \|\boldsymbol{w}_{c,i} - w^o\|^2, \tag{58}$$

$$\text{MSD}_{\text{avg},i} \triangleq \frac{1}{P}\sum_{p=1}^{P}\|\boldsymbol{w}_{p,i} - w^o\|^2. \tag{59}$$

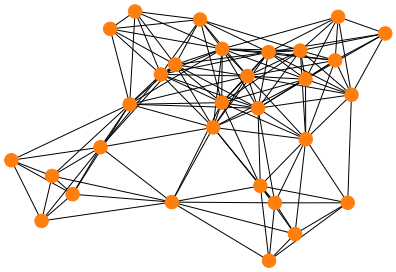


图3：生成的智能体网络。

如图4所示，与随机扰动不同，图同态扰动不会妨碍算法对真实模型的逼近性能。随机扰动的均方偏差曲线（黄色）明显高于图同态扰动的均方偏差曲线（红色），后者接近未受扰动的均方偏差曲线（蓝色）。考察个体模型的平均均方偏差时，我们发现其性能衰减程度不及随机扰动时严重。此外，由于局部图同态扰动不会影响算法的性能，我们观察到其均方偏差曲线与非隐私化算法的曲线一致。

B. 分布式学习中的分类

接下来，我们在一个分类数据集上进行实验。我们使用 Avazu 点击率数据集 [55], ，其中包含一组在线广告点击数据。我们将数据分配给 $P = 50$个智能体。为获得非独立同分布数据，我们向每个智能体的数据集中添加非独立同分布的高斯噪声。令 $\mu = 0.5$、 $\rho = 0.001$ 和 $\sigma_g^2 = 0.8$。我们在图5中绘制了标准算法、采用随机扰动的隐私化算法、采用图同态扰动的隐私化算法，以及采用局部图同态扰动的隐私化算法的测试误差。不出所料，与随机扰动一样，使用图同态扰动不会损害测试误差，而局部图同态扰动则不会改变测试误差。

V. 结论

本研究旨在探讨隐私对分布式学习的影响。我们证实，与随机扰动相比，图同态扰动在模型性能方面更具优势。随后，我们设计了局部图同态扰动，以确保添加噪声不会影响模型性能。因此，本研究的主要结论是，在分布式学习中，图同态扰动优于随机扰动。

附录A 分布式算法的敏感度

我们研究分布式学习算法 (21)–(23) 的敏感度，其在每个时刻由以下表达式定义：

$$\Delta(i) = \|\boldsymbol{\mathcal{W}}_i - \boldsymbol{\mathcal{W}}_i'\|. \tag{60}$$

此定义刻画了单个智能体的数据样本发生变化时所产生的变化。撇号表示新的轨迹。我们可以利用三角不等式，根据个体误差和最优模型之间的差异来界定敏感度：

$$\Delta(i) \le \|\widetilde{\boldsymbol{\mathcal{W}}}_i'\| + \|\widetilde{\boldsymbol{\mathcal{W}}}_i\| + \sqrt{P}\|w^o - w'^o\|. \tag{61}$$

然后，对于设计者选定的任意常数 B 和 B'，我们可以利用马尔可夫不等式得到以下界：

$$\mathbb{P}(\|\widetilde{\boldsymbol{\mathcal{W}}}_i\| \ge B) \le \frac{\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_i\|^2}{B^2}, \tag{62}$$

$$\mathbb{P}(\|\widetilde{\boldsymbol{\mathcal{W}}}_i'\| \ge B') \le \frac{\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_i'\|^2}{B'^2}. \tag{63}$$

现在回顾定理 1：

$$\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_i\|^2 \le \kappa_2^2\mathbb{1}^{\mathsf{T}}\Gamma^i\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0\|^2\\ \|\mathbb{E}\bar{\boldsymbol{\mathcal{W}}}_0\|^2\end{bmatrix} + O(\mu) + O(\mu^{-1}), \tag{64}$$

$$\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_i'\|^2 \le \kappa_2'^2\mathbb{1}^{\mathsf{T}}\Gamma'^i\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0'\|^2\\ \mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0'\|^2\end{bmatrix} + O(\mu) + O(\mu^{-1}). \tag{65}$$

由此可得，敏感度有如下界：

$$\Delta(i) \le B + B' + \sqrt{P}\|w^o - w'^o\| \tag{66}$$

以高概率成立，其中概率为：

$$\begin{aligned}&\mathbb{P}(\Delta(i) \le B + B' + \sqrt{P}\|w^o - w'^o\|)\\ &\ge \left(1 - \frac{\kappa_2^2\mathbb{1}^{\mathsf{T}}\Gamma^i\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0\|^2\\ \mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0\|^2\end{bmatrix} + O(\mu) + O(\mu^{-1})}{B^2}\right)\\ &\quad\times\left(1 - \frac{\kappa_2'^2\mathbb{1}^{\mathsf{T}}\Gamma'^i\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0'\|^2\\ \mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0'\|^2\end{bmatrix} + O(\mu) + O(\mu^{-1})}{B'^2}\right).\end{aligned} \tag{67}$$

Appendix B
Proof of Theorem 1

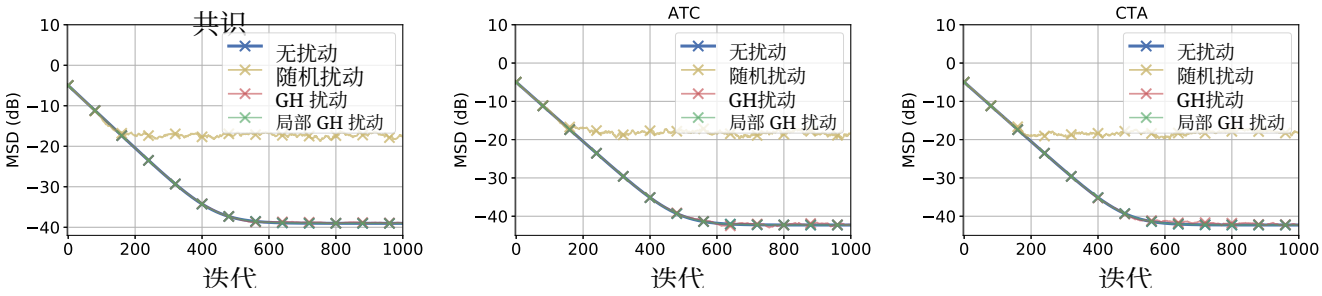
以下证明采用了与[26] 中针对非隐私算法的证明相似的步骤。利用 $A_2^{\mathsf{T}}A_0^{\mathsf{T}}A_1^{\mathsf{T}}$ 的约当分解：

$$A_2^{\mathsf{T}}A_0^{\mathsf{T}}A_1^{\mathsf{T}} = V_\theta J V_\theta^{-1}, \tag{68}$$

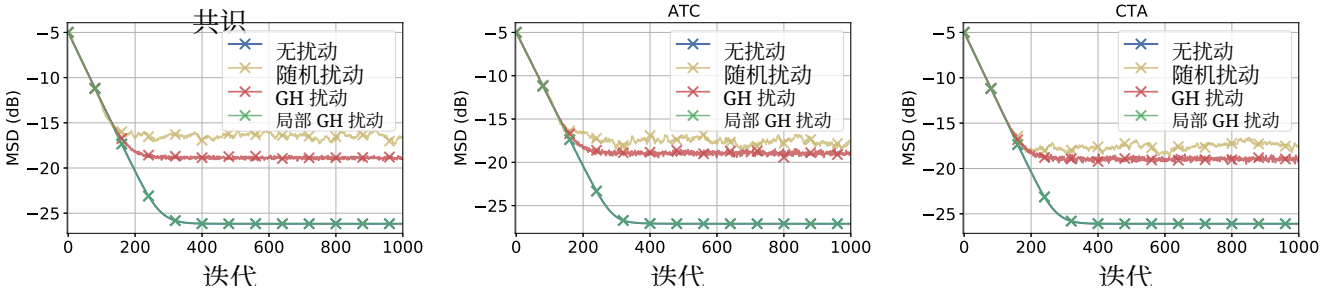
$$V_\theta \triangleq \begin{bmatrix}q & V_R\end{bmatrix}, \tag{69}$$

$$V_\theta^{-1} \triangleq \begin{bmatrix}\mathbb{1}^{\mathsf{T}}\\ V_L^{\mathsf{T}}\end{bmatrix}, \tag{70}$$

$$J \triangleq \begin{bmatrix}1 & 0\\ 0 & J_\theta\end{bmatrix}, \tag{71}$$

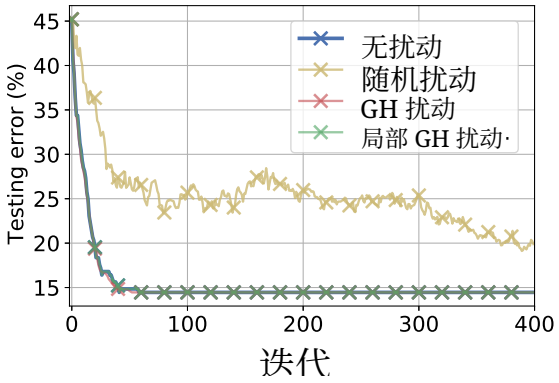


(a) 共识：质心 MSD (b) ATC：质心 MSD (c) CTA：质心 MSD

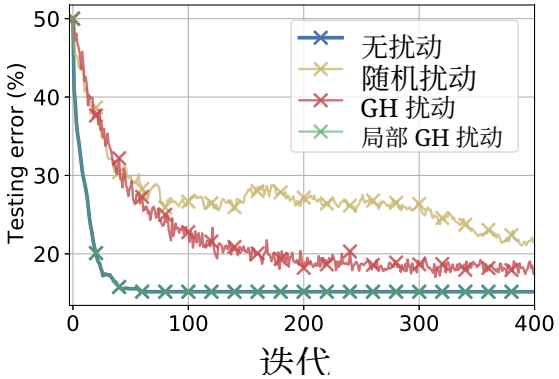


(d) 共识：平均个体 MSD (e) ATC：平均个体 MSD (f) CTA：平均个体 MSD

图4：三种分布式学习算法的均方偏差图。



(a) 质心测试误差



(b) 平均个体测试误差

图5：标准ATC、采用随机扰动的隐私化ATC，以及采用图同态扰动的隐私化ATC的测试误差。

其中， q 是 $A_2^{\mathsf{T}} A_0^{\mathsf{T}} A_1^{\mathsf{T}}$ 的佩龙特征向量，而 J_θ 包含对应特征值 λ 的约当块，其形式为

（ 3×3 矩阵示例）：

$$\begin{bmatrix} \lambda & 0 & 0 \\ \theta & \lambda & 0 \\ 0 & \theta & \lambda \end{bmatrix}, \tag{72}$$

w下副对角线上的常数为 θ 。我们首先 写作：

$$\boldsymbol{\mathcal{B}}_{i-1} = (\mathcal{V}_\theta^{-1})^{\mathsf{T}}(\mathcal{J} - \boldsymbol{\mathcal{D}}_{i-1}^{\mathsf{T}})\mathcal{V}_\theta^{\mathsf{T}}, \tag{73}$$

其中：

$$\mathcal{V}_\theta^{-1} \triangleq V_\theta^{-1} \otimes I_M, \tag{74}$$

$$\mathcal{V}_\theta \triangleq V_\theta \otimes I_M, \tag{75}$$

$$\mathcal{J} \triangleq J \otimes I_M = \begin{bmatrix} I_M & 0 \\ 0 & \mathcal{J}_\theta \end{bmatrix}, \tag{76}$$

$$\boldsymbol{\mathcal{D}}_{i-1}^{\mathsf{T}} \triangleq \mu \mathcal{V}_\theta^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \boldsymbol{\mathcal{H}}_{i-1} \mathcal{A}_1^{\mathsf{T}} (\mathcal{V}_\theta^{-1})^{\mathsf{T}} = \begin{bmatrix} \boldsymbol{D}_{11,i-1}^{\mathsf{T}} & \boldsymbol{D}_{21,i-1}^{\mathsf{T}} \\ \boldsymbol{D}_{12,i-1}^{\mathsf{T}} & \boldsymbol{D}_{22,i-1}^{\mathsf{T}} \end{bmatrix}. \tag{77}$$

在 [26] 中定理 9.1 的证明中已表明：

$$\|I_M - \boldsymbol{D}_{11,i-1}\| \le 1 - \sigma_{11}\mu, \tag{78}$$

$$\|\boldsymbol{D}_{ij}\| \le \sigma_{ij}\mu, \tag{79}$$

对于某些正常数 σ_{ij}，当 $i,\ j = 1,\ 2$。将误差递推 (41) 的两边左乘 $\mathcal{V}_\theta^{\mathsf{T}}$：

$$\begin{aligned} \mathcal{V}_\theta^{\mathsf{T}} \widetilde{\boldsymbol{\mathcal{W}}}_i = {} & \mathcal{V}_\theta^{\mathsf{T}} \boldsymbol{\mathcal{B}}_{i-1} (\mathcal{V}_\theta^{-1})^{\mathsf{T}} \mathcal{V}_\theta^{\mathsf{T}} \widetilde{\boldsymbol{\mathcal{W}}}_{i-1} + \mu \mathcal{V}_\theta^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \boldsymbol{\mathcal{S}}_i - \mu \mathcal{V}_\theta^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} b \\ & + \mu \mathcal{V}_\theta^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \boldsymbol{\mathcal{H}}_{i-1} \mathrm{diag}(\mathcal{A}_1^{\mathsf{T}} \boldsymbol{\mathcal{G}}_{1,i}) - \mathcal{V}_\theta^{\mathsf{T}} \mathrm{diag}(\mathcal{A}_2^{\mathsf{T}} \boldsymbol{\mathcal{G}}_{2,i}) \\ & - \mathcal{V}_\theta^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathrm{diag}(\mathcal{A}_0^{\mathsf{T}} \boldsymbol{\mathcal{G}}_{0,i}) - \mathcal{V}_\theta^{\mathsf{T}} \mathcal{A}_2^{\mathsf{T}} \mathcal{A}_0^{\mathsf{T}} \mathrm{diag}(\mathcal{A}_1^{\mathsf{T}} \boldsymbol{\mathcal{G}}_{1,i}), \end{aligned} \tag{80}$$

并引入新的记号：

$$\mathcal{V}_\theta^\mathsf{T}\widetilde{\boldsymbol{\mathcal{W}}}_i = \begin{bmatrix}(q^\mathsf{T}\otimes I_M)\widetilde{\boldsymbol{\mathcal{W}}}_i\\ (V_R^\mathsf{T}\otimes I)\widetilde{\boldsymbol{\mathcal{W}}}_i\end{bmatrix} \triangleq \begin{bmatrix}\bar{\boldsymbol{\mathcal{W}}}_i\\ \check{\boldsymbol{\mathcal{W}}}_i\end{bmatrix}, \tag{81}$$

$$\mu\mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}\boldsymbol{s}_i = \mu\begin{bmatrix}(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\boldsymbol{s}_i\\ (V_R^\mathsf{T}\otimes I)\mathcal{A}_2^\mathsf{T}\boldsymbol{s}_i\end{bmatrix} \triangleq \begin{bmatrix}\bar{\boldsymbol{s}}_i\\ \check{\boldsymbol{s}}_i\end{bmatrix}, \tag{82}$$

$$\mu\mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}b = \mu\begin{bmatrix}(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}b\\ (V_R^\mathsf{T}\otimes I)\mathcal{A}_2^\mathsf{T}b\end{bmatrix} \triangleq \begin{bmatrix}0\\ \check{b}\end{bmatrix}, \tag{83}$$

我们得到：

$$\begin{aligned}\begin{bmatrix}\bar{\boldsymbol{\mathcal{W}}}_i\\ \check{\boldsymbol{\mathcal{W}}}_i\end{bmatrix} =& \begin{bmatrix}I_M-\boldsymbol{D}_{11,i-1}^\mathsf{T} & -\boldsymbol{D}_{21,i-1}^\mathsf{T}\\ -\boldsymbol{D}_{12,i-1}^\mathsf{T} & \mathcal{J}_\epsilon\end{bmatrix}\begin{bmatrix}\bar{\boldsymbol{\mathcal{W}}}_{i-1}\\ \check{\boldsymbol{\mathcal{W}}}_{i-1}\end{bmatrix} + \begin{bmatrix}\bar{\boldsymbol{s}}_i\\ \check{\boldsymbol{s}}_i\end{bmatrix} + \begin{bmatrix}0\\ \check{b}\end{bmatrix}\\ &+\mu\mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{H}}_{i-1}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i}) - \mathcal{V}_\theta^\mathsf{T}\mathrm{diag}(\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i})\\ &-\mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i}) - \mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i}).\end{aligned} \tag{84}$$

然后，对 ℓ_2- 范数取期望，并利用詹森不等式，我们有：

$$\begin{aligned}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_i\|^2 \le& (1-\sigma_{11}\mu)\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_{i-1}\|^2 + \frac{\sigma_{21}^2\mu}{\sigma_{11}}\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_{i-1}\|^2 + \mathbb{E}\|\bar{\boldsymbol{s}}_i\|^2\\ &+2\mu^2\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{H}}_{i-1}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2+\\ &+2\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathrm{diag}(\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i})\|^2\\ &+\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i})\|^2,\end{aligned} \tag{85}$$

and:

$$\begin{aligned}\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_i\|^2 \le& \left(\rho(J_\theta)+\theta+\frac{2\sigma_{22}^2\mu^2}{1-\rho(J_\theta)-\theta}\right)\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_{i-1}\|^2\\ &+\frac{3\sigma_{21}^2\mu^2}{1-\rho(J_\theta)-\theta}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_{i-1}\|^2 + \frac{3\|\check{b}\|^2}{1-\rho(J_\theta)-\theta}\\ &+\mathbb{E}\|\check{\boldsymbol{s}}_i\|^2\\ &+2\mu^2\mathbb{E}\|(V_R^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{H}}_{i-1}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|J_\epsilon^\mathsf{T}(V_R^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|(V_R^\mathsf{T}\otimes I_M)\mathrm{diag}(\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i})\|^2\\ &+\mathbb{E}\|(V_R^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i})\|^2,\end{aligned} \tag{86}$$

由于零均值随机变量相互独立，交叉项为零。接着，我们对梯度噪声之和进行界定：

$$\begin{aligned}&\mathbb{E}\|\bar{\boldsymbol{s}}_i\|^2+\mathbb{E}\|\check{\boldsymbol{s}}_i\|^2\\ &\le \|\mathcal{V}_\theta\|^2\mu^2\sum_{p=1}^P\mathbb{E}\|\boldsymbol{s}_{p,i}\|^2\\ &\le \kappa_1^2\mu^2\sum_{p=1}^P\beta_{s,p}^2\mathbb{E}\|\widetilde{\boldsymbol{\phi}}_{p,i-1}\|^2+\sigma_{s,p}^2\\ &\le \kappa_1^2\mu^2\sum_{p=1}^P\beta_{s,p}^2\sum_{m=1}^P\left(\mathbb{E}\|\widetilde{\boldsymbol{w}}_{m,i-1}\|^2+\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2\right)+\sigma_{s,p}^2\\ &\le \kappa_1^2\mu^2\beta_s^2\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_{i-1}\|^2+\kappa_1^2\mu^2\sigma_s^2+\kappa_1^2\mu^2\sum_{p,m=1}^P\beta_{s,p}^2\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2\\ &\le \kappa_1^2\kappa_2^2\mu^2\beta_s^2\left(\mathbb{E}\|\bar{\boldsymbol{s}}_i\|^2+\mathbb{E}\|\check{\boldsymbol{s}}_i\|^2\right)+\kappa_1^2\mu^2\sigma_s^2\\ &\quad+\kappa_1^2\mu^2\sum_{p,m=1}^P\beta_{s,p}^2\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2,\end{aligned} \tag{87}$$

其中，我们引入了常数 β_s^2 和 σ_s^2，它们分别是 $\beta_{s,p}^2$ 和 $\sigma_{s,p}^2$ 之和。接着，回到前面的内容：

$$\begin{aligned}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_i\|^2 \le& (1-\sigma_{11}\mu+\kappa_1^2\kappa_2^2\beta_s^2\mu^2)\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_{i-1}\|^2\\ &+\left(\frac{\sigma_{21}^2\mu}{\sigma_{11}}+\kappa_1^2\kappa_2^2\mu^2\right)\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_{i-1}\|^2+\kappa_1^2\mu^2\sigma_s^2\\ &+\kappa_1^2\sum_{p,m=1}^P\beta_{s,p}^2\mu^2\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2\\ &+2\mu^2\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{H}}_{i-1}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+2\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathrm{diag}(\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i})\|^2\\ &+\mathbb{E}\|(q^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i})\|^2,\end{aligned} \tag{88}$$

and:

$$\begin{aligned}\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_i\|^2 \le& \left(\rho(J_\theta)+\theta+\frac{3\sigma_{22}^2\mu^2}{1-\rho(J_\theta)-\theta}+\kappa_1^2\kappa_2^2\beta_s^2\mu^2\right)\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_{i-1}\|^2\\ &+\left(\frac{3\sigma_{12}^2\mu^2}{1-\rho(J_\theta)-\theta}+\kappa_1^2\kappa_2^2\beta_s^2\mu^2\right)\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_{i-1}\|^2\\ &+\frac{3\|\check{b}\|^2}{1-\rho(J_\theta)-\theta}+\kappa_1^2\mu^2\sigma_s^2\\ &+\kappa_1^2\mu^2\sum_{p,m=1}^P\beta_{s,p}^2\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2\\ &+2\mu^2\mathbb{E}\|(V_R^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{H}}_{i-1}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|J_\theta^\mathsf{T}(V_R^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|(V_R^\mathsf{T}\otimes I_M)\mathrm{diag}(\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i})\|^2\\ &+\mathbb{E}\|(V_R^\mathsf{T}\otimes I_M)\mathcal{A}_2^\mathsf{T}\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i})\|^2.\end{aligned} \tag{89}$$

将两个界相加：

$$\begin{aligned}\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_i\|^2 \le& \kappa_2^2\Bigg(\bar{\gamma}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_{i-1}\|^2+\check{\gamma}\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_{i-1}\|^2+\frac{3\|\check{b}\|^2}{1-\rho(J_\theta)-\theta}\\ &+2\kappa_1^2\mu^2\sigma_s^2+2\kappa_1^2\mu^2\sum_{p,m=1}^P\beta_{s,p}^2\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2\\ &+2\mu^2\mathbb{E}\|\mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{H}}_{i-1}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|\mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}\mathcal{A}_0^\mathsf{T}\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\mathbb{E}\|\mathcal{V}_\theta^\mathsf{T}\mathrm{diag}\ (\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i})\|^2\\ &+\mathbb{E}\|\mathcal{V}_\theta^\mathsf{T}\mathcal{A}_2^\mathsf{T}\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i})\|^2\Bigg)\\ \le& \kappa_2^2\Bigg(\bar{\gamma}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_{i-1}\|^2+\check{\gamma}\mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_{i-1}\|^2+\frac{3\|\check{b}\|^2}{1-\rho(J_\theta)-\theta}\\ &+2\kappa_1^2\mu^2\sigma_s^2+2\kappa_1^2\mu^2\sum_{p,m=1}^P\beta_{s,p}^2\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2\\ &+\kappa_1^2(1+2\delta^2\mu^2)\mathbb{E}\|\mathrm{diag}(\mathcal{A}_1^\mathsf{T}\boldsymbol{\mathcal{G}}_{1,i})\|^2\\ &+\kappa_1^2\mathbb{E}\|\mathrm{diag}(\mathcal{A}_2^\mathsf{T}\boldsymbol{\mathcal{G}}_{2,i})\|^2\\ &+\kappa_1^2\mathbb{E}\|\mathrm{diag}(\mathcal{A}_0^\mathsf{T}\boldsymbol{\mathcal{G}}_{0,i})\|^2\Bigg).\end{aligned} \tag{90}$$

然后，递归地对均方误差取界，并令 i 趋于无穷大，得到：

$$\begin{aligned}
&\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_i\|^2\\
&\le \kappa_2^2\mathbb{1}^{\mathsf{T}}(I-\Gamma)^{-1}\\
&\times\begin{bmatrix}\kappa_1^2\mu^2\sigma_s^2+\left(4+(2\delta^2+P\beta_s^2)\mu^2\right)\sigma_g^2\\ \frac{3\|\check{b}\|^2}{1-\rho(J_\theta)-\theta}+\kappa_1^2\mu^2\sigma_s^2+\kappa_1^2\left(3+(2\delta^2+P\beta_s^2)\mu^2\right)\sigma_g^2\end{bmatrix}\\
&=O(\mu)\sigma_s^2+O(\mu)+(O(\mu^{-1})+O(\mu))\sigma_g^2, \qquad (91)
\end{aligned}$$

其中：

$$\Gamma\triangleq\begin{bmatrix}\bar{\gamma} & \frac{\sigma_{21}^2\mu}{\sigma_{11}}+\kappa_1^2\kappa_2^2\mu^2\\ \frac{3\sigma_{12}^2\mu^2}{1-\rho(J_\theta)-\theta}+\kappa_1^2\kappa_2^2\beta_s^2\mu^2 & \check{\gamma}\end{bmatrix}. \qquad (92)$$

附录 C 引理 1 的证明

我们定义 $\boldsymbol{\mathcal{W}}_{c,i}\triangleq(q\mathbb{1}^{\mathsf{T}}\otimes I)\boldsymbol{\mathcal{W}}_i$ 并写为：

$$\begin{aligned}
\boldsymbol{\mathcal{W}}_i-\boldsymbol{\mathcal{W}}_{c,i}&=\left(I-q\mathbb{1}^{\mathsf{T}}\otimes I\right)\boldsymbol{\mathcal{W}}_i\\
&=(V_L^{\mathsf{T}}\otimes I)(V_R\otimes I)\boldsymbol{\mathcal{W}}_i\\
&=(V_L^{\mathsf{T}}\otimes I)J_\theta(V_R\otimes I)\boldsymbol{\mathcal{W}}_{i-1}\\
&\quad-\mu(V_L^{\mathsf{T}}\otimes I)(V_R\otimes I)\mathrm{col}\left\{\nabla_{w^{\mathsf{T}}}J_p(\boldsymbol{\phi}_{p,i-1})\right\}\\
&\quad-\mu(V_L^{\mathsf{T}}\otimes I)(V_R\otimes I)\boldsymbol{s}_i\\
&\quad+(V_L^{\mathsf{T}}\otimes I)(V_R\otimes I)\big(\mathrm{diag}(\mathcal{A}_2^{\mathsf{T}}\boldsymbol{\mathcal{G}}_{2,i})\\
&\quad+\mathcal{A}_2^{\mathsf{T}}\mathrm{diag}(\mathcal{A}_0^{\mathsf{T}}\boldsymbol{\mathcal{G}}_{0,i})+\mathcal{A}_2^{\mathsf{T}}\mathcal{A}_0^{\mathsf{T}}\mathrm{diag}(\mathcal{A}_1^{\mathsf{T}}\boldsymbol{\mathcal{G}}_{1,i})\big). \qquad (93)
\end{aligned}$$

我们通过使用带有常数 $\rho(J_\theta)<1$ 的詹森不等式来界定 $\mathbb{E}\|(V_R\otimes I)\boldsymbol{\mathcal{W}}_i\|^2$，并定义 $\kappa_1^2=\|\mathcal{V}_\theta\|^2$ 和 $\kappa_2^2=\|\mathcal{V}_\theta^{-1}\|^2$：

$$\begin{aligned}
&\mathbb{E}\|(V_R\otimes I)\boldsymbol{\mathcal{W}}_i\|^2\\
&\le\rho(J_\theta)\mathbb{E}\|(V_R\otimes I)\boldsymbol{\mathcal{W}}_{i-1}\|^2\\
&\quad+\frac{\kappa_1^2\kappa_2^2\mu^2}{1-\rho(J_\theta)}\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{H}_{p,i-1}\widetilde{\boldsymbol{\phi}}_{p,i-1}+\nabla_{w^{\mathsf{T}}}J_p(w^o)\|^2\\
&\quad+\kappa_1^2\kappa_2^2\mu^2\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{s}_{p,i}\|^2+v_1^2v_2^2\big(\mathbb{E}\|\mathrm{diag}(\mathcal{A}_2^{\mathsf{T}}\boldsymbol{\mathcal{G}}_{2,i})\|^2\\
&\quad+\mathbb{E}\|\mathcal{A}_2^{\mathsf{T}}\mathrm{diag}(\mathcal{A}_0^{\mathsf{T}}\boldsymbol{\mathcal{G}}_{0,i})\|^2+\mathbb{E}\|\mathcal{A}_2^{\mathsf{T}}\mathcal{A}_0^{\mathsf{T}}\mathrm{diag}(\mathcal{A}_1^{\mathsf{T}}\boldsymbol{\mathcal{G}}_{1,i})\|^2\big)\\
&\le\rho(J_\theta)\mathbb{E}\|(V_R\otimes I)\boldsymbol{\mathcal{W}}_{i-1}\|^2+\frac{2\kappa_1^2\kappa_2^2\mu^2\|b\|^2}{1-\rho(J_\theta)}\\
&\quad+\kappa_1^2\kappa_2^2\mu^2\sum_{p=1}^{P}\beta_{s,p}^2\sum_{m\in\mathcal{N}_p}a_{1,mp}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{m,i-1}\|^2\\
&\quad+\kappa_1^2\kappa_2^2\mu^2\sum_{p,m=1}^{P}a_{1,mp}^2\sigma_g^2+\kappa_1^2\kappa_2^2\mu^2\sigma_s^2\\
&\quad+\kappa_1^2\kappa_2^2\sum_{p,m=1}^{P}(a_{2,mp}^2+a_{0,mp}^2+a_{1,mp}^2)\sigma_g^2. \qquad (94)
\end{aligned}$$

然后，个体误差 $\mathbb{E}\|\widetilde{\boldsymbol{w}}_{m,\,i-1}\|^2$ 可按照定理 1 进行界定：

$$\begin{aligned}
\mathbb{E}\|\widetilde{\boldsymbol{w}}_{m,i-1}\|^2&\le\mathbb{E}\|\widetilde{\boldsymbol{\mathcal{W}}}_{i-1}\|^2\\
&\le\kappa_2^2\mathbb{1}^{\mathsf{T}}\left(\Gamma^{i-1}\begin{bmatrix}\mathbb{E}\|\bar{\boldsymbol{\mathcal{W}}}_0\|^2\\ \mathbb{E}\|\check{\boldsymbol{\mathcal{W}}}_0\|^2\end{bmatrix}\right.\\
&\quad\left.+(I-\Gamma)^{-1}(I-\Gamma^{i-1})\begin{bmatrix}O(\mu^2)+O(1)\\ O(\mu^2)+O(1)\end{bmatrix}\right), \qquad (95)
\end{aligned}$$

其中，$O(\mu^2)$ 和 $O(1)$ 项是依赖于梯度噪声方差、偏差项 b 和噪声方差的常数。此外，矩阵 Γ 刻画了递推的收敛速率，并已在附录 B 中定义。

然后，我们将此界代回主不等式 (94)，并对 i 进行递归界定。网络分歧的界为：

$$\frac{1}{P}\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{w}_{p,i}-\boldsymbol{w}_{c,i}\|^2\le\frac{\kappa_2^2}{P}\mathbb{E}\|(V_R\otimes I)\boldsymbol{\mathcal{W}}_i\|^2, \qquad (96)$$

并且在极限情况下：

$$\begin{aligned}
&\limsup_{i\to\infty}\frac{1}{P}\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{w}_{p,i}-\boldsymbol{w}_{c,i}\|^2\\
&\le\frac{2\kappa_1^2\kappa_2^4\mu^2\|b\|^2}{P(1-\rho(J_\theta))^2}+\frac{\kappa_1^2\kappa_2^4}{P(1-\rho(J_\theta))}\sigma_g^2+O(\mu)\\
&\quad+\frac{\kappa_1^2\kappa_2^4}{P(1-\rho(J_\theta))}\left(O(\mu^2)\sigma_s^2+O(\mu^2)\sigma_g^2\right). \qquad (97)
\end{aligned}$$

附录 D 定理 2 的证明

从（46）出发，对过去的模型的欧几里得范数平方取条件均值后，我们可以将该范数拆分为三个相互独立的项：模型误差、梯度噪声和添加噪声。再次取期望，并结合 $\alpha=\sqrt{1-2\nu\mu+\delta^2\mu^2}$ 使用 Jensen 不等式，可得：

$$\begin{aligned}
&\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i}\|^2\\
&\le\alpha\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i-1}\|^2+\mu^2\mathbb{E}\left\|(q^{\mathsf{T}}\otimes I)\boldsymbol{s}_i\right\|^2\\
&\quad+\frac{\mu^2}{1-\alpha}\mathbb{E}\left\|\sum_{p=1}^{P}q_p\boldsymbol{H}_{p,i-1}\sum_{m\in\mathcal{N}_p}a_{1,mp}(\boldsymbol{w}_{m,i-1}-\boldsymbol{w}_{c,i-1})\right\|^2\\
&\quad+\mu^2\mathbb{E}\left\|\sum_{p=1}^{P}q_p\boldsymbol{H}_{p,i-1}\sum_{m\in\mathcal{N}_p}a_{1,mp}\boldsymbol{g}_{1,mp,i}\right\|^2\\
&\le\alpha\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i-1}\|^2+\mu^2\mathbb{E}\left\|(q^{\mathsf{T}}\otimes I)\boldsymbol{s}_i\right\|^2\\
&\quad+\frac{\delta^2\mu^2}{1-\alpha}\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{w}_{p,i-1}-\boldsymbol{w}_{c,i-1}\|^2\\
&\quad+\mu^2\mathbb{E}\left\|\sum_{p=1}^{P}q_p\boldsymbol{H}_{p,i-1}\sum_{m\in\mathcal{N}_p}a_{1,mp}\boldsymbol{g}_{1,mp,i}\right\|^2. \qquad (98)
\end{aligned}$$

我们从（34）出发，并使用 Jensen 不等式引入 $\widetilde{\boldsymbol{w}}_{c,\,i-1}$，以此界定梯度噪声：

$$
\begin{aligned}
&\mathbb{E}\left\|(q^{\mathsf{T}}\otimes I)\boldsymbol{s}_i^2\right\|^2\\
&=\sum_{p=1}^{P}q_p^2\mathbb{E}\|\boldsymbol{s}_{p,i}\|^2\\
&\le\sum_{p=1}^{P}q_p^2\beta_{s,p}^2\mathbb{E}\|\widetilde{\boldsymbol{\phi}}_{p,i-1}\|^2+q_p^2\sigma_{s,p}^2\\
&\le\sum_{p=1}^{P}q_p^2\beta_{s,p}^2\sum_{m\in\mathcal{N}_p}a_{1,mp}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{m,i-1}\|^2+a_{1,mp}\mathbb{E}\|\boldsymbol{g}_{1,mp,i}\|^2\\
&\quad+\sigma_s^2\\
&\le 2\beta_s^2\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i-1}\|^2+\sigma_s^2+\beta_s^2\sigma_g^2\\
&\quad+2\sum_{p=1}^{P}q_p^2\beta_{s,p}^2\sum_{m\in\mathcal{N}_p}a_{1,mp}\mathbb{E}\|\boldsymbol{w}_{m,i-1}-\boldsymbol{w}_{c,i-1}\|^2\\
&\le 2\beta_s^2\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i-1}\|^2+\sigma_s^2+\beta_s^2\sigma_g^2\\
&\quad+2\beta_s^2\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{w}_{p,i-1}-\boldsymbol{w}_{c,i-1}\|^2.
\end{aligned}
\tag{99}
$$

使用两次詹森不等式，可以将噪声项界定如下：

$$
\mathbb{E}\left\|\sum_{p=1}^{P}q_p\boldsymbol{H}_{p,i-1}\sum_{m\in\mathcal{N}_p}a_{1,mp}\boldsymbol{g}_{1,mp,i}\right\|^2\le\delta^2\sigma_g^2. \tag{100}
$$

将梯度噪声和添加的隐私噪声的界代入(98)：

$$
\begin{aligned}
\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i}\|^2\le\left(\alpha+2\beta_s^2\mu^2\right)\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i-1}\|^2+\mu^2\sigma_s^2+(\beta_s^2+\delta^2)\mu^2\sigma_g^2\\
+\left(2\beta_s^2+\frac{\delta^2}{1-\alpha}\right)\mu^2\sum_{p=1}^{P}\mathbb{E}\|\boldsymbol{w}_{p,i-1}-\boldsymbol{w}_{c,i-1}\|^2.
\end{aligned}
\tag{101}
$$

我们使用引理 1 中的界。递归地界定误差的二阶矩并取极限：

$$
\begin{aligned}
\limsup_{i\to\infty}\mathbb{E}\|\widetilde{\boldsymbol{w}}_{c,i}\|^2&\le\frac{\mu^2\left(\sigma_s^2+(\beta_s^2+\delta^2)\sigma_g^2\right)}{1-\gamma_c}+O(\mu^2)\\
&\quad+\frac{\mu}{1-\gamma_c}\left(2\beta_s^2+\frac{\delta^2}{1-\alpha}\right)\frac{\kappa_1^2\kappa_2^4}{1-\rho(J_\theta)}\sigma_g^2\\
&=O(\mu)\sigma_s^2+O(1)\sigma_g^2+O(\mu^2).
\end{aligned}
\tag{102}
$$

附录E 定理3的证明

只需证明局部图同态过程生成的噪声服从拉普拉斯分布，因为我们已经知道，添加拉普拉斯噪声可以使算法以高概率满足差分隐私（见 [36]，[46]）。众所周知，均匀随机变量$U(0，1)$与伽马随机变量 $\Gamma(2，1)$的乘积服从指数随机变量 Exp(1) [56]。于是，$e^{-\boldsymbol{v}_\ell\boldsymbol{v}_m}$ 在 $[0,1]$ 上服从均匀分布：

$$
\mathbb{P}(e^{-\boldsymbol{v}_\ell\boldsymbol{v}_m}\le c)=\mathbb{P}(\boldsymbol{v}_\ell\boldsymbol{v}_m\ge-\ln c)=e^{\ln c}=c. \tag{103}
$$

但将它乘以 a 后，所得随机变量便在 $[0,a]$ 上服从均匀分布。只要 a 是 π 的倍数，均匀随机变量对 p 取模后仍在 $[0,\pi]$ 上服从均匀分布。令 $a=t\pi$，其中 t 为整数，且 $\boldsymbol{x}\sim\mathrm{U}(0，a)$。我们将区间 $[0,a]$ 划分为 t 个互不相交、长度为 π 的子区间，$[0,a]=[0,1)\cup[1,2)\cdots\cup[(t-1)\pi，a]$。在这些子区间中的每一个上，$[i\pi,(i+1)\pi)$ 都服从均匀分布，$\boldsymbol{x}$ 也如此$\mathbb{P}\big(\boldsymbol{x}\le x|\boldsymbol{x}\in[i\pi,(i+1)\pi)\big)=x$，因此 $\boldsymbol{x}$ 对 $\pi=\boldsymbol{x}-\lfloor\boldsymbol{x}/\pi\rfloor=\boldsymbol{x}-i\pi$ 取模后，在 $[0,\pi]$ 上服从均匀分布。由此，既然 $a=t\pi$，我们得到：

$$
\begin{aligned}
\mathbb{P}(\boldsymbol{x}\le x)&=\sum_{i=0}^{t-1}\mathbb{P}\big(\boldsymbol{x}\le x|\boldsymbol{x}\in[i\pi,(i+1)\pi)\big)\\
&\qquad\times\mathbb{P}\big(\boldsymbol{x}\in[i\pi,(i+1)\pi)\big)\\
&=\sum_{i=0}^{t-1}x\frac{\pi}{a}=x.
\end{aligned}
\tag{104}
$$

这意味着 $\boldsymbol{v}_{\ell m}\sim\mathrm{U}(0，\pi)$。然后，两个指数随机变量之差服从拉普拉斯分布。因此，我们需要将两个均匀随机变量转换为参数为$\frac{\sigma_g}{\sqrt{2}}$的两个指数随机变量。取 $-\frac{\sqrt{2}}{\sigma_g}\ln\boldsymbol{v}_{\ell m}$即可得到一个指数随机变量：

$$
\mathbb{P}\left(-\frac{\sqrt{2}}{\sigma_g}\ln(\boldsymbol{v}_{\ell m})\le c\right)=\mathbb{P}\left(\boldsymbol{v}_{\ell m}\ge e^{-\frac{c\sigma_g}{\sqrt{2}}}\right)=1-e^{-\frac{\sigma_g c}{\sqrt{2}}}. \tag{105}
$$

参考文献

[1] D. Bertsekas，《最小二乘问题的一类新型增量梯度方法》，*SIAM*优化期刊，第7卷，第913–926页，1997年11月。[2] D. Blatt、A. O. Hero 和 H. Gauchman，《一种具有恒定步长的收敛增量梯度方法》，*SIAM*优化期刊，第18卷，第29–51页，2007年。[3] F. S. Cattivelli 和 A. H. Sayed，《分布式估计中的空间与增量最小均方处理分析》，*IEEE*信号处理汇刊，第59卷，第4期，第1465–1480页，2011年。[4] C. G. Lopes 和 A. H. Sayed，《分布式网络上的增量自适应策略》，*IEEE*信号处理汇刊，第55卷，第8期，第4064–4077页，2007年。[5] B. Johansson、M. Rabi 和 M. Johansson，《一种用于网络化系统分布式优化的随机增量次梯度方法》，*SIAM*优化期刊，第20卷，第1157–1170页，2009年1月。[6] E. S. Helou 和 A. R. De Pierro，《约束凸优化的增量次梯度：统一框架与新方法》，*SIAM*优化期刊，第20卷，第3期，第1547–1572页，2009年12月。[7] M. G. Rabbat 和 R. D. Nowak，《用于分布式优化的量化增量算法》，*IEEE*通信精选领域期刊，第23卷，第4期，第798–808页，2005年。[8] A. Nedic 和 D. P. Bertsekas，《不可微优化的增量次梯度方法》，*SIAM*优化期刊，第12卷，第1期，第109–138页，2001年。[9] M. H. DeGroot，《达成共识》，美国统计学会会刊。，第69卷，第345期，第118–121页，1974年。[10] A. Nedic 和 A. Ozdaglar，《多智能体优化的分布式次梯度方法》，*IEEE*自动控制汇刊，第54卷，第1期，第48–61页，2009年。[11] L. Xiao 和 S. Boyd，《用于分布式平均的快速线性迭代》，系统与控制快报，第53卷，第1期，第65–78页，2004年9月。[12] S. Barbarossa 和 G. Scutari，《仿生传感器网络设计》，*IEEE* 信号处理杂志，第 24 卷，第3期，第26–35页，2007年5月。[13] B. Johansson、T. Keviczky、M. Johansson 和 K. H. Johansson，《求解凸优化问题的次梯度方法与共识算法》，载于*IEEE*控制决策会议（*CDC*）论文集，墨西哥坎昆，2008年12月，第4185–4190页。[14] W. Ren 和 R. Beard，《动态变化交互拓扑下多智能体系统的共识寻求》，*IEEE*自动控制汇刊，第50卷，第5期，第655–661页，2005年。

[15] S. Boyd、A. Ghosh、B. Prabhakar 和 D. Shah，《随机化闲聊算法》，IEEE信息论汇刊，第52卷，第6期，第2508–2530页，2006年6月。[16] S. Kar 和 J. M. F. Moura，《不完善通信条件下传感器网络中的分布式共识算法：链路故障与信道噪声》，IEEE信号处理汇刊，第57卷，第1期，第355–369页，2009年1月。[17] K. Srivastava 和 A. Nedic，《分布式异步约束随机优化》，IEEE信号处理精选主题期刊，第5卷，第4期，第772–790页，2011年8月。[18] O. Hlinka、O. Slu˘ciak、F. Hlawatsch、P. M. Djuri´c 和 M. Rupp，《似然共识及其在分布式粒子滤波中的应用》，IEEE信号处理汇刊，第60卷，第8期，第4334–4349页，2012年8月。[19] J. Chen 和 A. H. Sayed，《用于网络分布式优化与学习的扩散自适应策略》，IEEE信号处理汇刊，第60卷，第8期，第4289–4305页，2012年8月。[20] S.-Y. Tu 和 A. H. Sayed，《对于自适应网络中的分布式估计，扩散策略优于共识策略》，IEEE信号处理汇刊，第60卷，第12期，第6217–6234页，2012年12月。[21] J. Chen 和 A. H. Sayed，《自适应网络的学习行为——第一部分：瞬态分析》，IEEE信息论汇刊，第61卷，第6期，第3487–3517页，2015年12月。[22] S. Chouvardas、K. Slavakis、Y. Kopsinis 和 S. Theodoridis，《一种用于分布式学习的稀疏性促进自适应算法》，IEEE信号处理汇刊，第60卷，第10期，第5412–5425页，2012年。[23] S. Vlaski 和 A. H. Sayed，《非凸环境中的分布式学习——第一部分：线性速率一致性》，IEEE信号处理汇刊，第69卷，第1242–1256页，2021年。[24] C. G. Lopes 和 A. H. Sayed，《自适应网络上的扩散最小均方：公式与性能分析》，IEEE信号处理汇刊，第56卷，第7期，第3122–3136页，2008年7月。[25] A. H. Sayed，《自适应网络》，IEEE会刊，第102卷，第4期，第460–497页，2014年。[26] ——，《网络上的自适应、学习与优化》，《机器学习基础与趋势》，第7卷，801，2014 4 5, 311 no. - pp. – 。[27] B. Hitaj、G. Ateniese 和 F. Perez-Cruz，《生成对抗网络下的深度模型：协作式深度学习中的信息泄露》，载于ACM SIGSAC计算机与通信安全会议论文集，美国纽约，2017年，第603–618页。[28] L. Melis、C. Song、E. De Cristofaro 和 V. Shmatikov，《利用协作式学习中的非预期特征泄露》，载于IEEE安全与隐私研讨会，美国加利福尼亚州旧金山，2019年5月，第691–706页。[29] M. Nasr、R. Shokri 和 A. Houmansadr，《深度学习的全面隐私分析：针对集中式学习和联邦学习的被动与主动白盒推理攻击》，载于IEEE安全与隐私研讨会，美国加利福尼亚州圣何塞，2019年5月，第739–753页。[30] L. Zhu 和 S. Han，《梯度深度泄露》，载于神经信息处理系统进展，加拿大温哥华，2019年12月，第17–31页。[31] K. Bonawitz、V. Ivanov、B. Kreuter、A. Marcedone、H. B. McMahan、S. Patel、D. Ramage、A. Segal 和 K. Seth，《用于隐私保护机器学习的实用安全聚合》，载于ACM SIGSAC计算机与通信安全会议论文集，美国纽约，2017年，第1175–1191页。[32] A. Gasc´on、P. Schoppmann、B. Balle、M. Raykova、J. Doerner、S. Zahur 和 D. Evans，《高维数据的隐私保护分布式线性回归》，隐私增强技术论文集，第2017卷，第4期，第345–364页，2017年。[33] P. Mohassel 和 Y. Zhang，《Secureml：一种可扩展隐私保护机器学习系统》，载于IEEE安全与隐私研讨会，美国加利福尼亚州圣何塞，2017年5月，第19–38页。[34] D. Froelicher、J. R. Troncoso-Pastoriza、A. Pyrgelis、S. Sav、J. S. Sousa、J.-P. Bossuat 和 J.-P. Hubaux，《可扩展隐私保护分布式学习》，隐私增强技术论文集，第2021卷，第2期，第323–347页，2021年。[35] V. Nikolaenko、U. Weinsberg、S. Ioannidis、M. Joye、D. Boneh 和 N. Taft，《数亿条记录的隐私保护岭回归》，载于IEEE安全与隐私研讨会，美国加利福尼亚州伯克利，2013年5月，第334–348页。[36] C. Dwork 和 A. Roth，《差分隐私的算法基础》。Found. Trends Theor. Comput. Sci.，第9卷，第3–4期，第211–407页，2014年。[37] R. C. Geyer、T. Klein 和 M. Nabi，“差分隐私联邦学习：客户端层面的视角”，arXiv:1712.07557，2017年。[38] R. Hu、Y. Guo、H. Li、Q. Pei 和 Y. Gong，“差分隐私个性化联邦学习”，IEEE物联网期刊，第7卷，第10期，第9530–9539页，2020年。[39] A. Triastcyn 和 B. Faltings，“采用贝叶斯差分隐私的联邦学习”，载于 IEEE大数据国际会议，洛杉矶（加利福尼亚州），2019年12月，第2587–2596页。[40] S. Truex、L. Liu、K.-H. Chow、M. E. Gursoy 和 W. Wei，“Ldp-fed：采用本地差分隐私的联邦学习”，载于第三届 ACM 边缘系统、分析与网络国际研讨会论文集，伊拉克利翁（希腊），2020年4月，第61–66页。[41] K. Wei、J. Li、M. Ding、C. Ma、H. H. Yang、F. Farokhi、S. Jin、T. Q. Quek 和 H. V. Poor，“差分隐私联邦学习：算法与性能分析”，IEEE信息取证与安全汇刊，第15卷，第3454–3469页，2020年。[42] B. Jayaraman、L. Wang、D. Evans 和 Q. Gu，“无忧分布式学习：隐私保护经验风险最小化”，载于《神经信息处理系统进展》，第31卷，蒙特利尔（加拿大），2018年12月。[43] C. Li、P. Zhou、L. Xiong、Q. Wang 和 T. Wang，“差分隐私分布式在线学习”，IEEE知识与数据工程汇刊，第30卷，第8期，第1440–1453页，2018年。[44] J. Zhu、C. Xu、J. Guan 和 D. O. Wu，“时变有向网络上的差分隐私分布式在线算法”，IEEE信号与信息处理网络汇刊，第4卷，第1期，第4–17页，2018年。[45] M. A. Pathak、S. Rane 和 B. Raj，“通过聚合本地训练的分类器实现多方差分隐私”，载于《神经信息处理系统进展》，温哥华（加拿大），2010年12月，第1876–1884页。[46] S. Vlaski 和 A. H. Sayed，“用于隐私保护去中心化学习的图同态扰动”，载于 ICASSP，多伦多（加拿大），2021年6月，第5240–5244页。[47] B. Ying 和 A. H. Sayed，“随机次梯度学习的性能极限，第二部分：多智能体情形”，《信号处理》，第144卷，第253–264页，2018年。[48] S. Vlaski 和 A. H. Sayed，“非凸环境中的分布式学习——第一部分：线性速率一致性”，IEEE信号处理汇刊，第69卷，第1242–1256页，2021年。[49] A. H. Sayed，《从数据中推断与学习》。剑桥大学出版社，2023年。[50] R. Nassif、S. Vlaski、C. Richard、J. Chen 和 A. H. Sayed，“图上的多任务学习：一种分布式流式机器学习方法”，IEEE信号处理杂志，第37卷，第25页，2020年3月，第14期。[51] S. Wang、T. Tuor、T. Salonidis、K. K. Leung、C. Makaya、T. He 和 K. Chan，“资源受限边缘计算系统中的自适应联邦学习”，IEEE通信领域精选期刊，第37卷，第6期，第1205–1221页，2019年。[52] X. Li、K. Huang、W. Yang、S. Wang 和 Z. Zhang，“论 FedAvg 在非独立同分布数据上的收敛性”，载于国际学习表征会议，亚的斯亚贝巴（埃塞俄比亚），2020年，第1–26页。[53] Ginar，“区块链上的随机数生成器（RNG）综述”，2019年12月。[在线]。网址：https://medium.com/ginar-io/
a-
r
eview-of-rando
m-number-generator-rng-on-blockchain-fe342d76261b [54] Y. Wang 和 H. V. Poor，“具有内在隐私保护的去中心化随机优化”，IEEE自动控制汇刊，第1–16页，2022年。[55] Avazu 和 Kaggle，“Avazu 的点击率预测”，2014年。[在线]。网址：http://www.csie.ntu.edu.tw/-cj1in/libsvmtools/ [56] N. L. Johnson、S. Kotz 和 N. Balakrishnan，《连续单变量分布》。Wiley，1995年。